

Machine Learning: Overview

UPDATE: The conda environments are now supported under an Anaconda site license. In this article, the list of available environments has been updated to support this change. General usage of the software has not changed; however, there are changes to the procedures for [cloning conda environments](#) and [activating an environment using c-shell \(csh\)](#).

Machine learning, deep learning, and artificial intelligence have become essential tools for handling and gaining insight from the enormous amounts of data that are being generated via high-performance computing, modern modeling and simulation, and instrument technology.

These methods organize data in a way that is both *systematic* (collected, processed, and stored methodically according to a standard practice) and *semantic* (unambiguous and logical, in both order and representation, so that relationships and biases can be easily explored). Once the data is organized, it can then be analyzed using powerful statistical methods.

HECC Data Science Team: Pilot Projects

HECC offers a range of services for researchers moving into advanced data science, using the latest technologies in statistics-based data analytics, machine learning, and deep learning. For general information, see [HECC Data Science Services](#).

The Data Science Team is currently working with NASA researchers on a number of projects, including:

- [Carbon Nanotube Gas Sensor Using Neural Networks](#) (PDF)
- [Predicting Composition of Photo Voltaic Cells Using Neural Networks](#) (PDF)

For more information about working with our team, please contact us at dataanalytics@nas.nasa.gov.

Getting Started with Machine Learning

In this section of the Knowledge Base, we provide information that can help you get started with a machine learning environment and short tutorials describing how to use the machine learning tools available on NAS systems. Please check back often, as new articles will become available soon.

Once you have a [NAS account](#), you will have access to software and infrastructure for your machine learning project. If you are a new user, be sure to read through our [New User Orientation](#) and complete the [first-time login process](#).

Conda Environments

We provide multiple conda environments that include basic machine learning packages, as well as common image processing and natural language processing packages, for your machine learning projects.

The following table shows currently available conda environments (all include GPU support):

Environment Name	Packages Included	Not Compatible
------------------	-------------------	----------------

			With
horovod	jupyterlab mpi4py pandas pillow	horovod tensorflow 2.4.1 pytorch 1.7.1	
pyt1_8	dask jupyterlab matplotlib mpi4py pandas	pillow pytorch 1.8.0+cu102 pytorch lightning 1.3.3 scikit-learn	
pyt1_10_2	dask jupyterlab matplotlib mpi4py pandas	pillow pytorch 1.10.2 pytorch lightning 1.5.9 scikit-learn	Sandy Bridge GPUs
pyt1_11	dask jupyterlab matplotlib mpi4py pandas	pillow pytorch 1.10.2 pytorch lightning 1.6 scikit-learn	Sandy Bridge GPUs
tf1_15	dask jupyterlab matplotlib nlTK pandas	pillow scikit-learn sympy tensorflow 1.15	
tf2_8	dask gensym jupyterlab matplotlib nlTK opencv	pandas pillow scikit-learn sympy tensorflow 2.8 tf_agents	
r3_6	jupyterlab r-caret r-dplyr r-essentials r-ggplot2 r-knitr r-markdown r-plyr	r-randomforest r-reticulate r-shiny r-tidyr r-tidyverse r-timedate r-xts r-zoo	
jupyterlab	jupyterlab		

The environments contain many of the same software programs, but the package versions may be different. Also, the environments may contain additional packages to the ones listed here. To learn more, see [Using Conda Environments for Machine Learning](#).

The conda environments are provided via software modules in HECC system directories. For general information about NAS-provided software modules, see [Software Directories](#) and [Using Software Modules](#).

Note: Rendering graphics using OpenAI Gym is currently unavailable.

New Packages

Miniconda has been upgraded to version 4.12.0 and is now licensed from Anaconda. The new license directly impacts how environments can be cloned. Some of the previous cloning options cannot be used with the licensed Anaconda environments. For the new cloning instructions, see [Managing and Installing Python Packages for Machine Learning](#).

Singularity Images

We also provide a few pre-built Singularity images containing popular machine learning packages, built for use on the GPU nodes. The images are installed in the `/swbuild/analytix/singularity/images` directory. You can use the images as provided, or you can copy them to your `/nobackup` directory to modify them. To learn how to use these images, see [Running a Singularity Container Image on Pleiades](#).

The following table shows currently available Singularity images:

Image File	Packages Included
tf2_4.sif	python 3.8 tensorflow 2.4 jupyterlab nltk
tf1_15.sif	python 3.8 tensorflow 1.15 jupyterlab matplotlib nltk numba pandas pillow scikit-learn
pyt1_8.sif	python 3.8 pytorch 1.8 jupyterlab matplotlib nltk numba pillow scikit-learn

Working with PBS

All NAS systems utilize the [Portable Batch System \(PBS\)](#) for batch job submission, job monitoring, and job management. To request examples of PBS scripts for Apache Spark and TensorFlow, please contact us at dataanalytics@nas.nasa.gov.

System Information

- [Systems and Filesystems](#)
- [GPU nodes](#)
- [Current system status](#)

Note: When you publish your results, please acknowledge the use of NAS resources.

Additional Resources

- HECC Webinars, including:
 - ◆ 2021 NAS Data Science Webinar
 - ◆ HECC Data Science Platform
- Machine Learning (Wikipedia)
- Machine Learning for Humans

Article ID: 572

Last updated: 14 Jun, 2022

Revision: 89

Machine Learning -> Machine Learning at NAS -> Machine Learning: Overview

<https://www.nas.nasa.gov/hecc/support/kb/entry/572/>